NEURAL INFORMATION PROCESSING SYSTEMS

Characteristic Circuits

Zhongjie Yu

TU Darmstadt
Germany

Martin Trapp

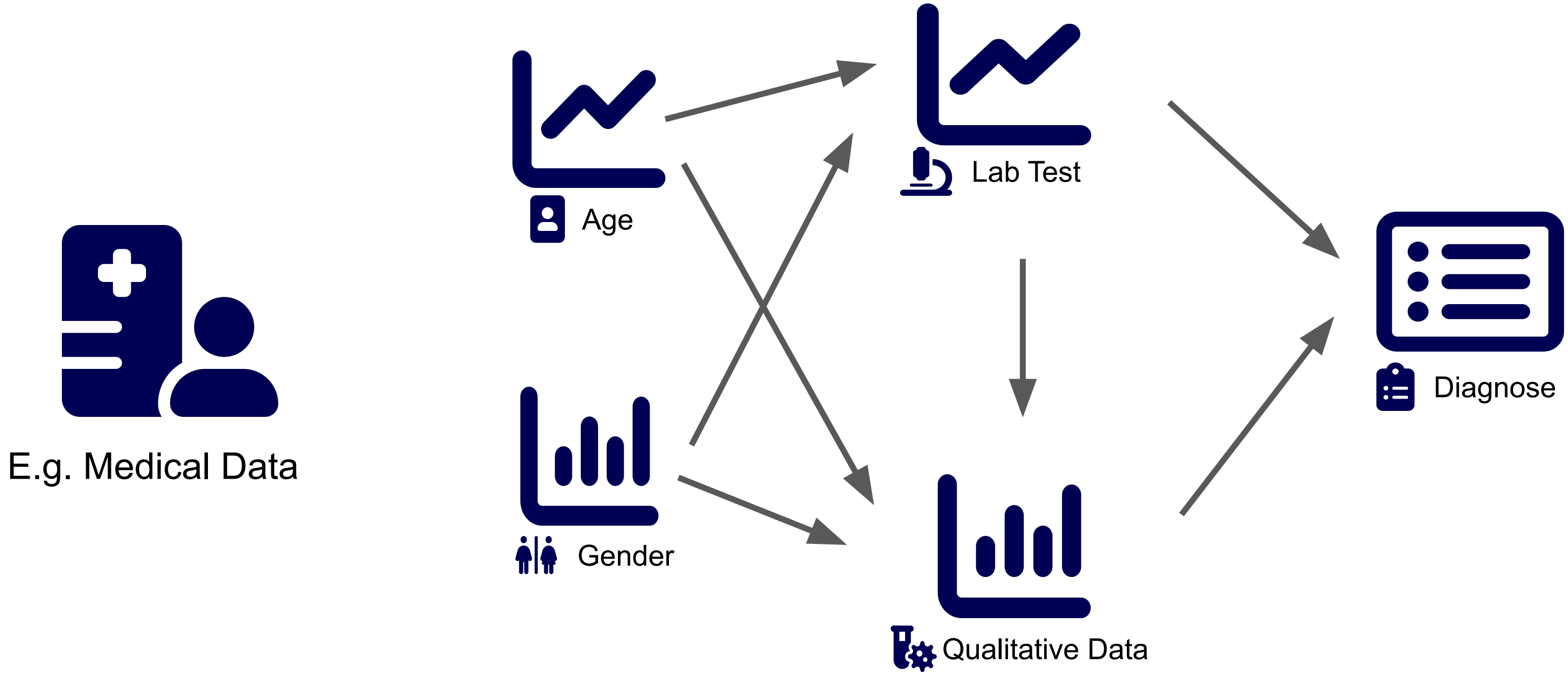
Aalto University
Finland

Kristian Kersting

TU Darmstadt/Hessian.AI/DFKI
Germany

Bundesministerium für Bildung und Forschung
Kompetenzzentren Arbeitsforschung
KompAKI
hessian.AI
DFKI
TECHNISCHE UNIVERSITÄT DARMSTADT
Aalto University School of Science

Real-world Data is often Hybrid

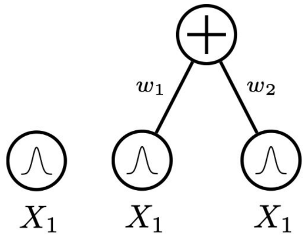


Probabilistic Circuits in a Nutshell

$$p(X_1)$$

[1] Figure from: Antonio Vergari et al., "*Tractable Probabilistic Models*", Talk at UAI Tutorial, 2019.
[2] YooJung Choi et al., "*Probabilistic circuits: A unifying framework for tractable probabilistic models*", Technical report, UCLA, 2020.

Probabilistic Circuits in a Nutshell

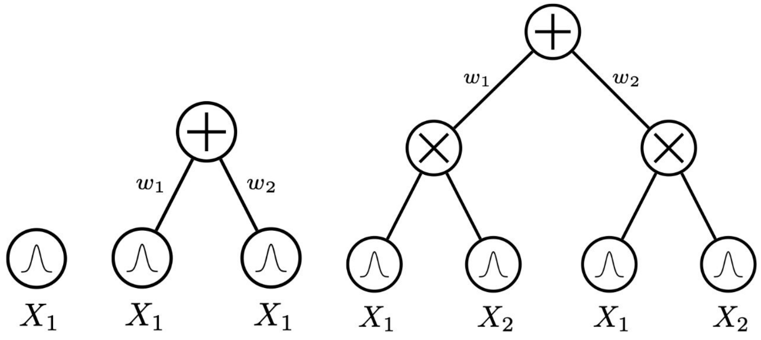


$p(X_1)$ $p(X_1)$

mixture

[1] Figure from: Antonio Vergari et al., "*Tractable Probabilistic Models*", Talk at UAI Tutorial, 2019.
[2] YooJung Choi et al., "*Probabilistic circuits: A unifying framework for tractable probabilistic models*", Technical report, UCLA, 2020.

Probabilistic Circuits in a Nutshell

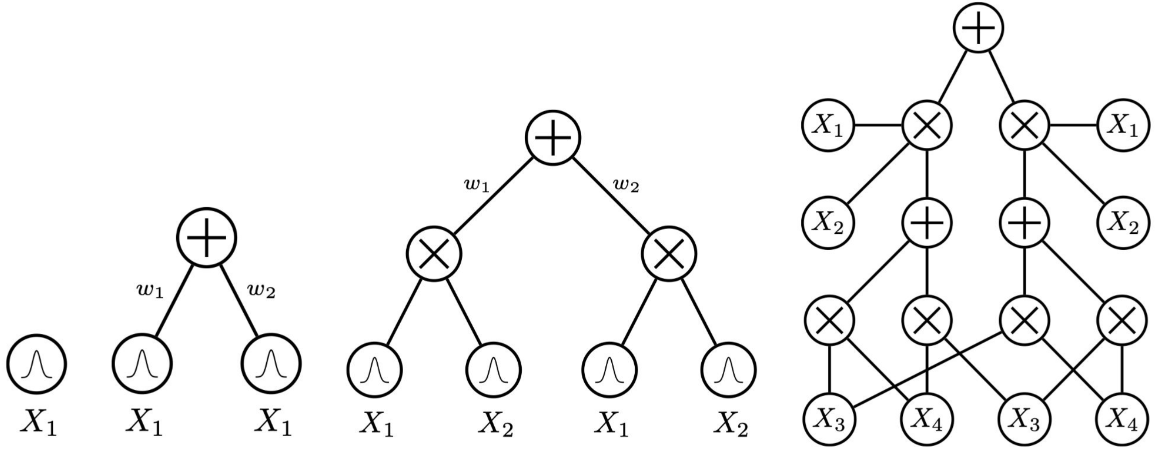


$p(X_1)$ $p(X_1)$ $p(X_1, X_2)$

mixture independence

[1] Figure from: Antonio Vergari et al., "*Tractable Probabilistic Models*", Talk at UAI Tutorial, 2019.
[2] YooJung Choi et al., "*Probabilistic circuits: A unifying framework for tractable probabilistic models*", Technical report, UCLA, 2020.

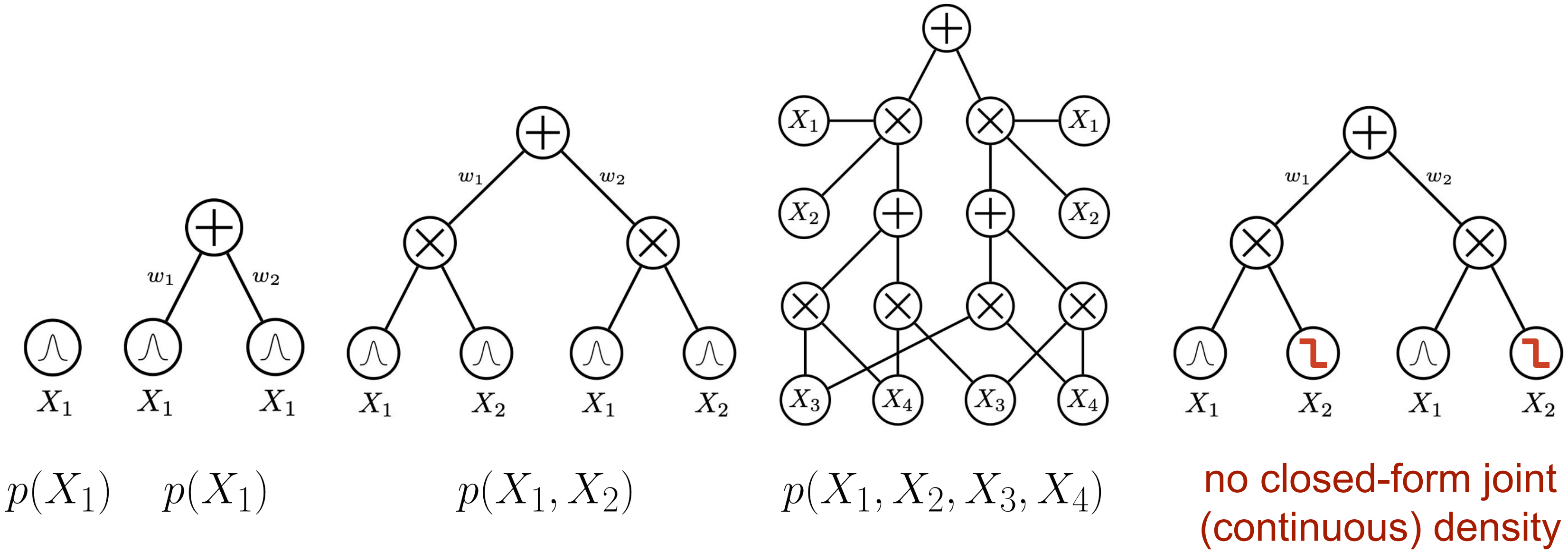
Probabilistic Circuits in a Nutshell



$p(X_1)$ $p(X_1)$ $p(X_1, X_2)$ $p(X_1, X_2, X_3, X_4)$

[1] Figure from: Antonio Vergari et al., "*Tractable Probabilistic Models*", Talk at UAI Tutorial, 2019.
[2] YooJung Choi et al., "*Probabilistic circuits: A unifying framework for tractable probabilistic models*", Technical report, UCLA, 2020.

Probabilistic Circuits in a Nutshell



[1] Figure from: Antonio Vergari et al., "*Tractable Probabilistic Models*", Talk at UAI Tutorial, 2019.
[2] YooJung Choi et al., "*Probabilistic circuits: A unifying framework for tractable probabilistic models*", Technical report, UCLA, 2020.

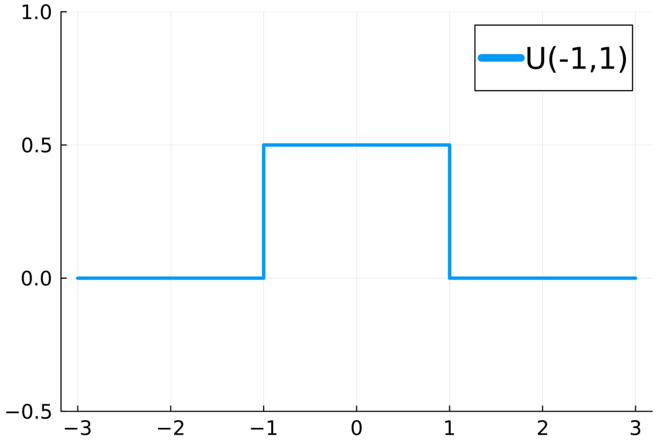
Characteristic Circuits in a Nutshell

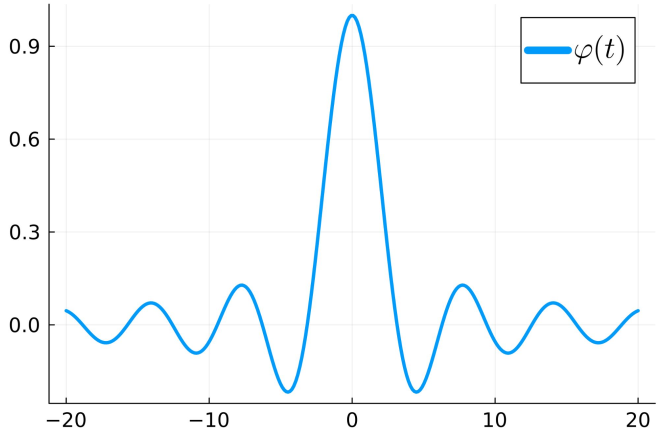
In mixed domains,
don’t go for histograms,
don’t go for PDFs,
don’t go for CDFs,
go for Characteristic Functions!

Characteristic Functions via the Fourier Transform

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}[\exp(\mathrm{i}\,\boldsymbol{t}^\top \boldsymbol{X})] = \int_{\boldsymbol{x}\in\mathbb{R}^d} \exp(\mathrm{i}\,\boldsymbol{t}^\top \boldsymbol{x})\,\mu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x})$$

Uniform Distribution



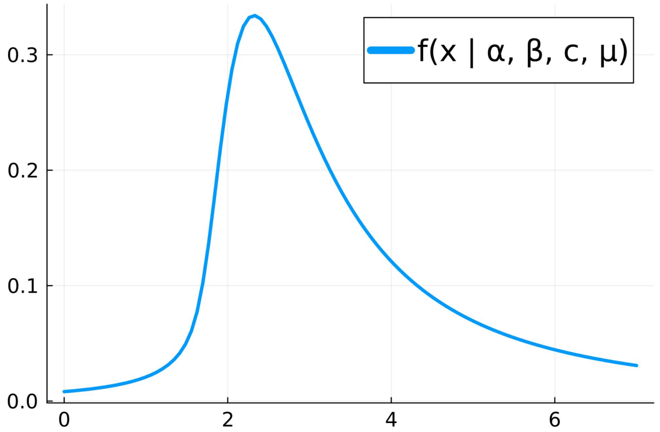


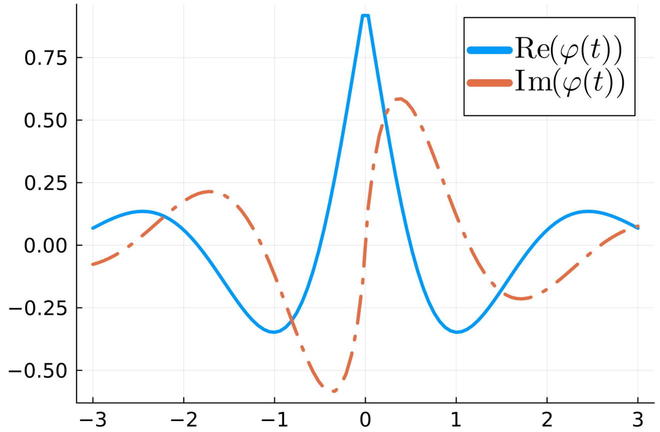
[1] Zoltán Sasvári, “*Multivariate characteristic and correlation functions*”, volume 50. Walter de Gruyter, 2013.
[2] John P Nolan, “*Multivariate elliptically contoured stable distributions: theory and estimation*”, Computational statistics, 2013.

Characteristic Functions via the Fourier Transform

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}[\exp(\mathrm{i}\,\boldsymbol{t}^\top \boldsymbol{X})] = \int_{\boldsymbol{x}\in\mathbb{R}^d} \exp(\mathrm{i}\,\boldsymbol{t}^\top \boldsymbol{x})\,\mu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x})$$

Alpha-Stable Distribution





[1] Zoltán Sasvári, “*Multivariate characteristic and correlation functions*”, volume 50. Walter de Gruyter, 2013.
[2] John P Nolan, “*Multivariate elliptically contoured stable distributions: theory and estimation*”, Computational statistics, 2013.

TECHNISCHE UNIVERSITÄT DARMSTADT

Aalto University School of Science

NEURAL INFORMATION PROCESSING SYSTEMS

Characteristic Functions via the Fourier Transform

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}[\exp(\mathrm{i}\,\boldsymbol{t}^\top \boldsymbol{X})] = \int_{\boldsymbol{x}\in\mathbb{R}^d} \exp(\mathrm{i}\,\boldsymbol{t}^\top \boldsymbol{x})\,\mu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x})$$

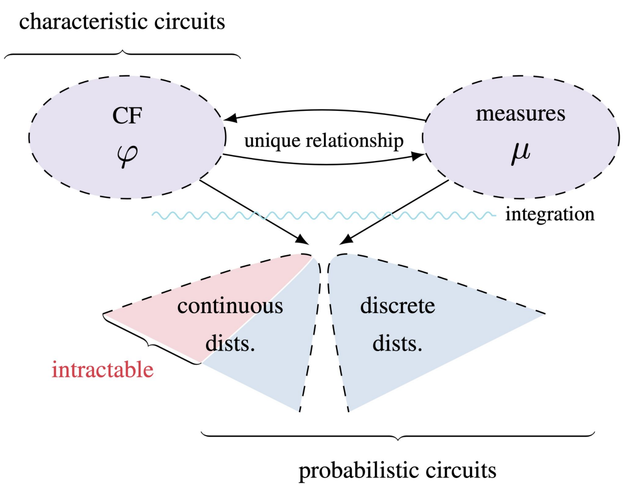
Categorical Distribution

$$P(X = j) = p_j \qquad \varphi_X(t) = \sum_{j=1}^{k} p_j \exp(\mathrm{i}\,t\,j)$$

[1] Zoltán Sasvári, "*Multivariate characteristic and correlation functions*", volume 50. Walter de Gruyter, 2013.
[2] John P Nolan, "*Multivariate elliptically contoured stable distributions: theory and estimation*", Computational statistics, 2013.

Characteristic Circuits are ...

- a tractable probabilistic model with
- a unified formalization of distributions over heterogeneous data
- in the continuous spectral domain with
- efficient probabilistic inference even when no closed-form density function is available



TECHNISCHE UNIVERSITÄT DARMSTADT
Aalto University School of Science

Characteristic Circuit - A Recursive Definition

- a ***characteristic function*** for a scalar random variable is a *characteristic circuit*

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}[\exp(\mathrm{i}\,\boldsymbol{t}^\top \boldsymbol{X})] = \int_{\boldsymbol{x}\in\mathbb{R}^d} \exp(\mathrm{i}\,\boldsymbol{t}^\top \boldsymbol{x})\,\mu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x})$$

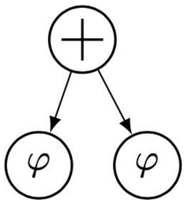
Characteristic Circuit - A Recursive Definition

- a ***characteristic function*** for a scalar random variable is a *characteristic circuit*

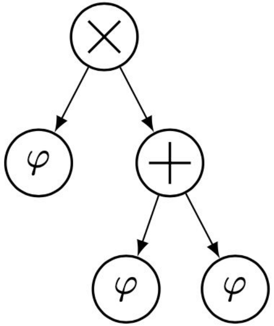
$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}[\exp(\mathrm{i}\,\boldsymbol{t}^\top \boldsymbol{X})] = \int_{\boldsymbol{x}\in\mathbb{R}^d} \exp(\mathrm{i}\,\boldsymbol{t}^\top \boldsymbol{x})\,\mu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x})$$

- a ***convex combination*** of characteristic circuits is a *characteristic circuit*

$$\int_{\boldsymbol{x}\in\mathbb{R}^d} \exp(\mathrm{i}\boldsymbol{t}^\top \boldsymbol{x}) \left[\sum\nolimits_{\mathsf{N}\in\mathrm{ch}(\mathsf{S})} w_{\mathsf{S},\mathsf{N}}\,\mu_{\mathsf{N}}(\mathrm{d}\boldsymbol{x})\right] = \sum\nolimits_{\mathsf{N}\in\mathrm{ch}(\mathsf{S})} w_{\mathsf{S},\mathsf{N}} \underbrace{\int_{\boldsymbol{x}\in\mathbb{R}^{p_{\mathsf{S}}}} \exp(\mathrm{i}\boldsymbol{t}^\top \boldsymbol{x})\,\mu_{\mathsf{N}}(\mathrm{d}\boldsymbol{x})}_{=\varphi_{\mathsf{N}}(\boldsymbol{t})}$$



Characteristic Circuit - A Recursive Definition



- a ***characteristic function*** for a scalar random variable is a *characteristic circuit*

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}[\exp(\mathrm{i}\,\boldsymbol{t}^\top \boldsymbol{X})] = \int_{\boldsymbol{x}\in\mathbb{R}^d} \exp(\mathrm{i}\,\boldsymbol{t}^\top \boldsymbol{x})\,\mu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x})$$

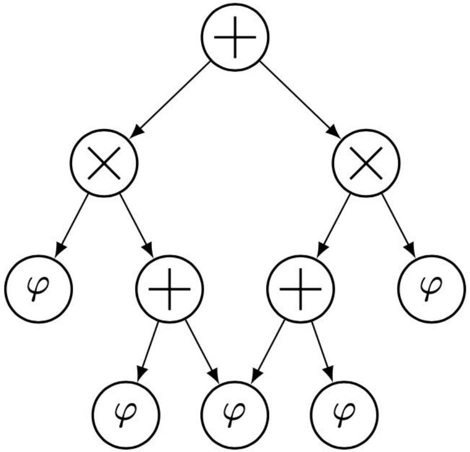
- a ***convex combination*** of characteristic circuits is a *characteristic circuit*

$$\int_{\boldsymbol{x}\in\mathbb{R}^d} \exp(\mathrm{i}\boldsymbol{t}^\top \boldsymbol{x}) \left[\sum\nolimits_{\mathsf{N}\in\mathrm{ch}(\mathsf{S})} w_{\mathsf{S},\mathsf{N}}\,\mu_{\mathsf{N}}(\mathrm{d}\boldsymbol{x})\right] = \sum\nolimits_{\mathsf{N}\in\mathrm{ch}(\mathsf{S})} w_{\mathsf{S},\mathsf{N}} \underbrace{\int_{\boldsymbol{x}\in\mathbb{R}^{p_\mathsf{S}}} \exp(\mathrm{i}\boldsymbol{t}^\top \boldsymbol{x})\,\mu_{\mathsf{N}}(\mathrm{d}\boldsymbol{x})}_{=\varphi_{\mathsf{N}}(\boldsymbol{t})}$$

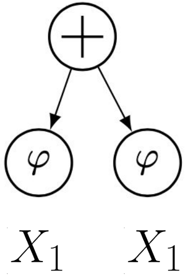
- a ***product*** of characteristic circuits is a *characteristic circuit*

$$\varphi_{\mathsf{P}}(\boldsymbol{t}) = \prod\nolimits_{\mathsf{N}\in\mathrm{ch}(\mathsf{P})} \varphi_{\mathsf{N}}(\boldsymbol{t}_{\psi(\mathsf{N})})$$

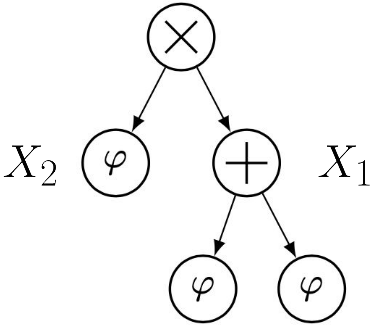
Characteristic Circuit - A Recursive Definition



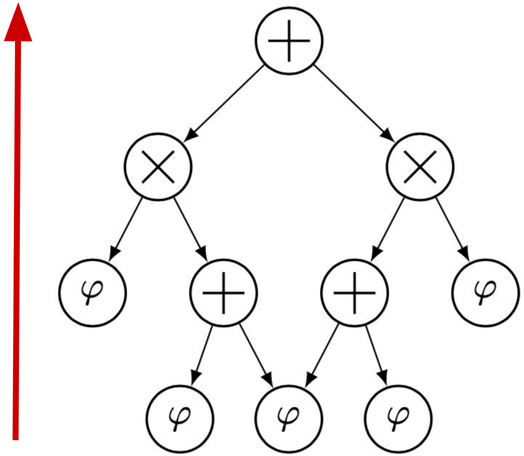
Smoothness:



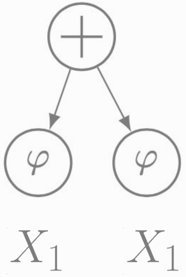
Decomposability:



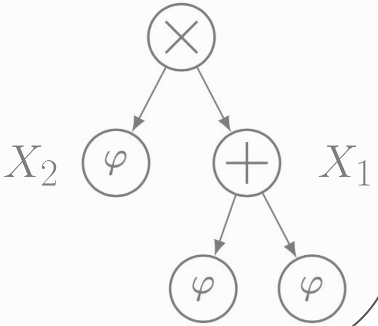
Characteristic Circuit - A Recursive Definition



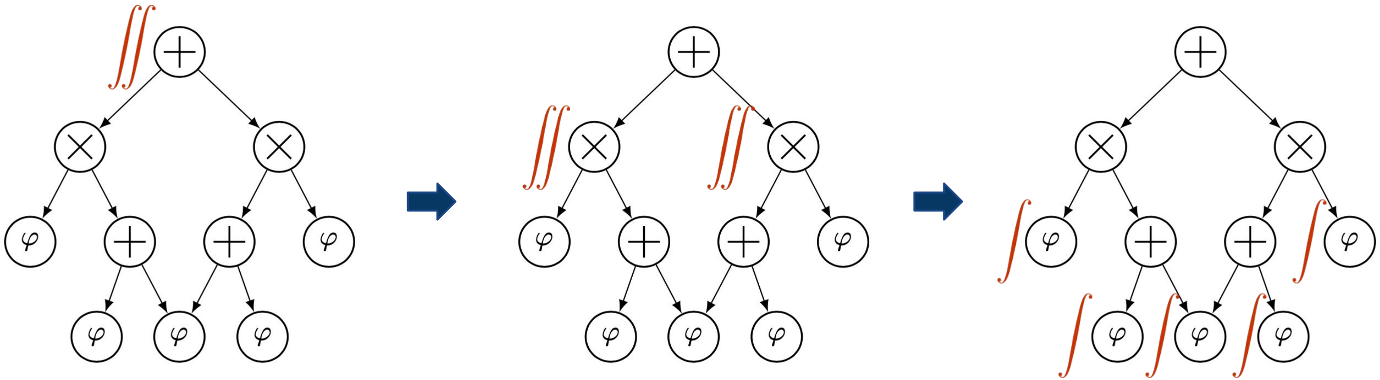
Smoothness:



Decomposability:

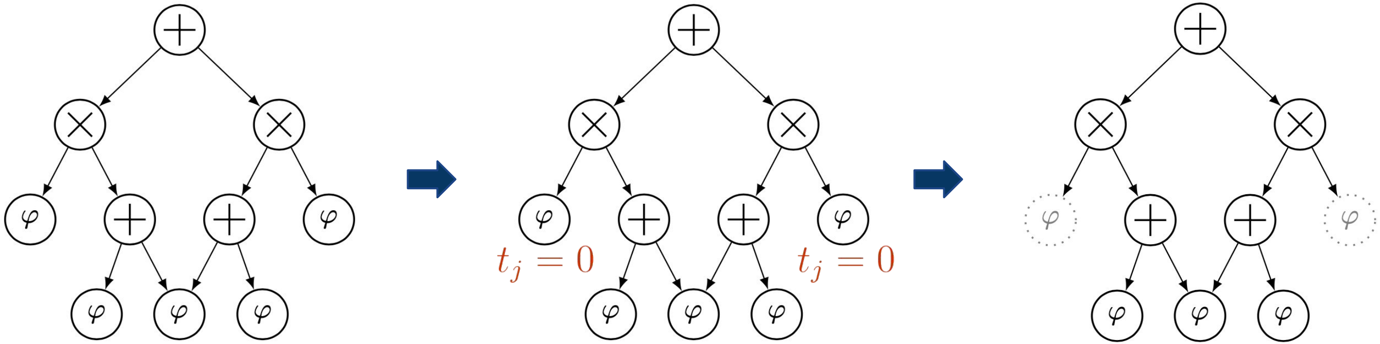


Efficient Computation of … Densities



Efficient computation through analytic or numerical integration at the leaves.

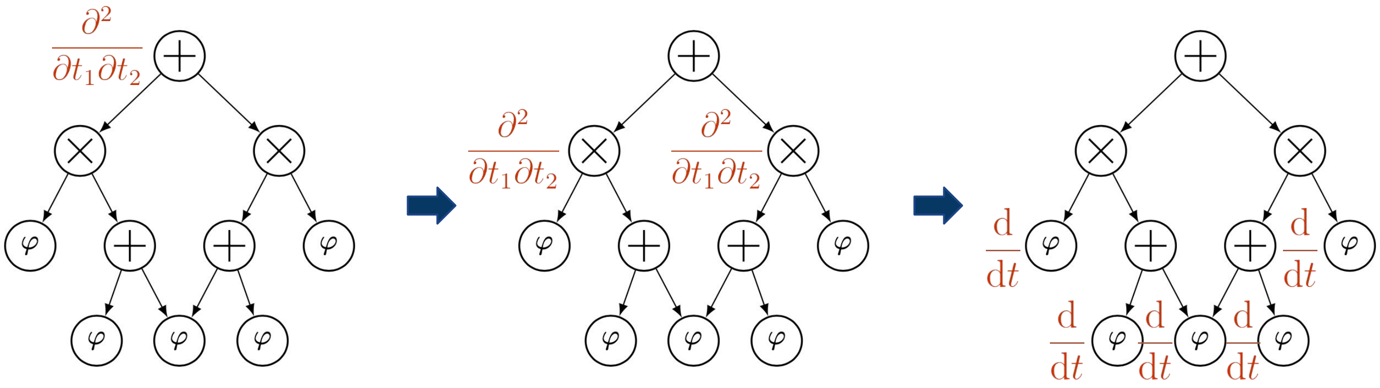
Efficient Computation of … Marginals



The marginal CC is obtained by setting the corresponding $t_j = 0$.

TECHNISCHE UNIVERSITÄT DARMSTADT
Aalto University School of Science

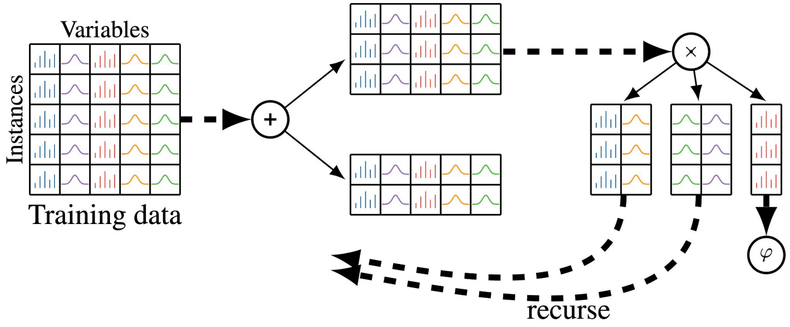
Efficient Computation of … Moments



The moments can be computed efficiently through differentiation of the circuit.

Learning Characteristic Circuits

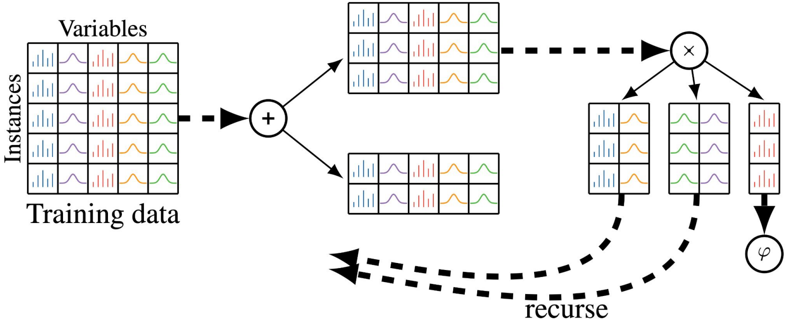
- Structure Learning - inspired by the LearnSPN algorithm



[1] Robert Gens and Pedro Domingos, "*Learning the structure of sum-product networks*", ICML, 2013.

Learning Characteristic Circuits

- Structure Learning - inspired by the LearnSPN algorithm



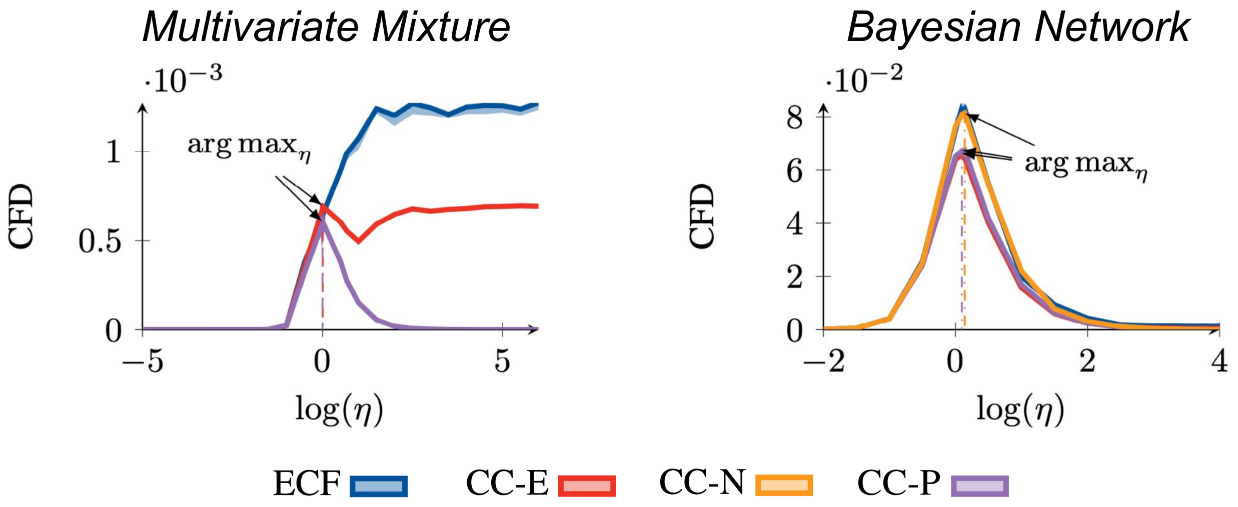
- Parameter Learning - minimizing the squared Characteristic Function Distance (CFD)

$$\frac{1}{k}\sum_{j=1}^{k}\left|\varphi_{\mathrm{ECF}}(\boldsymbol{t}_j) - \varphi_{\mathcal{C}}(\boldsymbol{t}_j)\right|^2 \quad \text{where} \quad \{t_1, \cdots, t_k\} \overset{\text{i.i.d.}}{\sim} \omega(\boldsymbol{t};\eta)$$

[1] Robert Gens and Pedro Domingos, "*Learning the structure of sum-product networks*", ICML, 2013.
[2] Abdul Fatir Ansari et al., "*A characteristic function approach to deep implicit generative modeling*", CVPR, 2020.

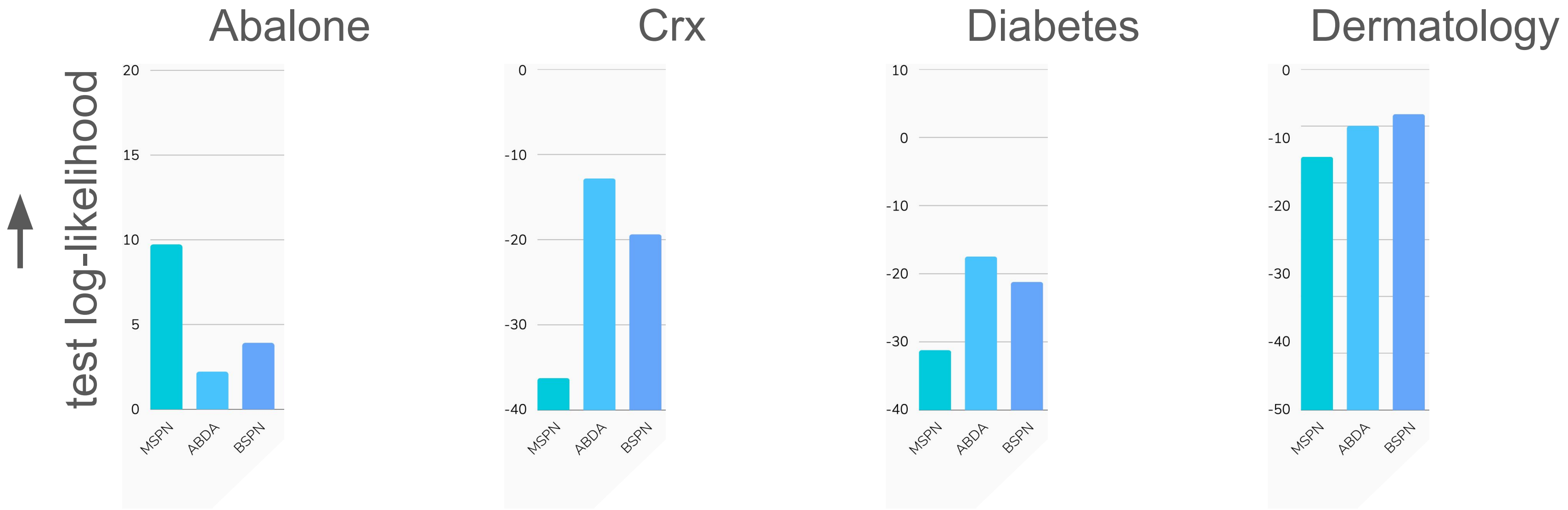
Characteristic Circuits Approximate Distributions Well

Compared to the widely used Empirical Characteristic Function (ECF, $\frac{1}{n}\sum_{j=1}^{n}\exp(\mathrm{i}\,t\,x_j)$)



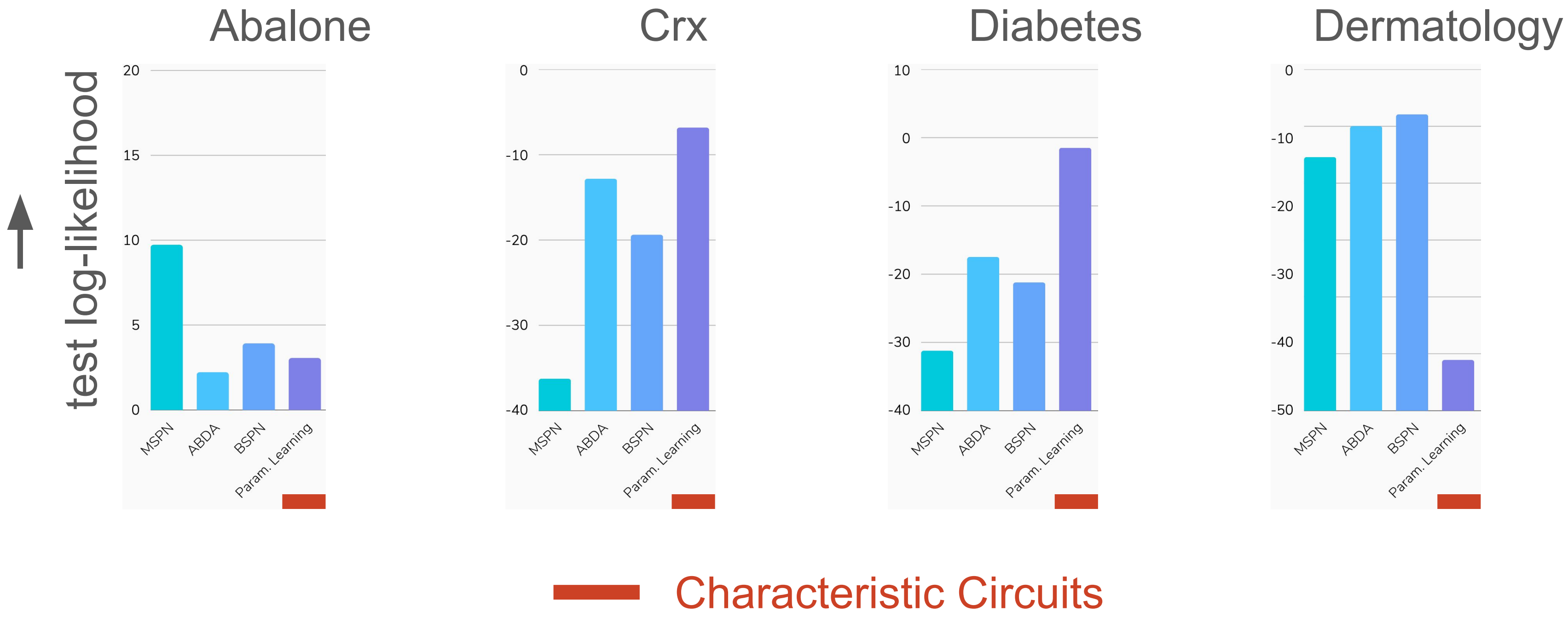
Discrete RV	ECF	ECF	Gaussian	Categorical
Continuous RV		ECF	Gaussian	Gaussian

Better Density Estimators on Heterogeneous Data

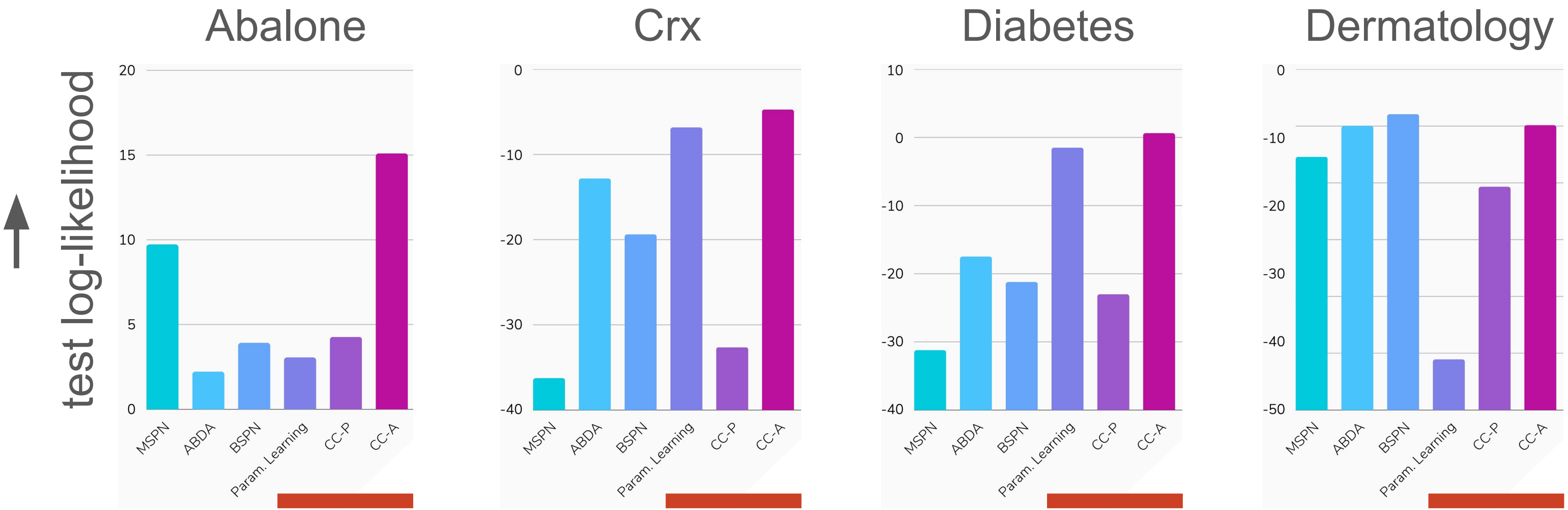


[1] Alejandro Molina et al., "*Mixed sum-product networks: A deep architecture for hybrid domains*", AAAI, 2018.
[2] Antonio Vergari et al., "*Automatic bayesian density analysis*", AAAI, 2019.
[3] Martin Trapp et al., "*Bayesian learning of sum-product networks*", NeurIPS, 2019.

Better Density Estimators on Heterogeneous Data

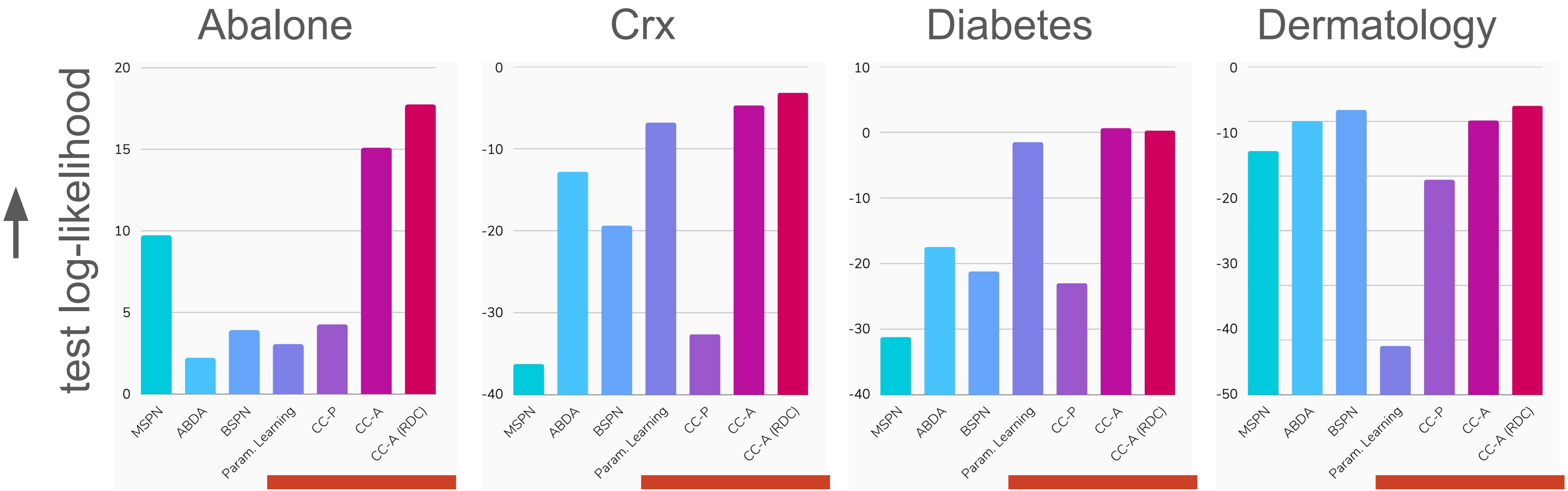


Better Density Estimators on Heterogeneous Data



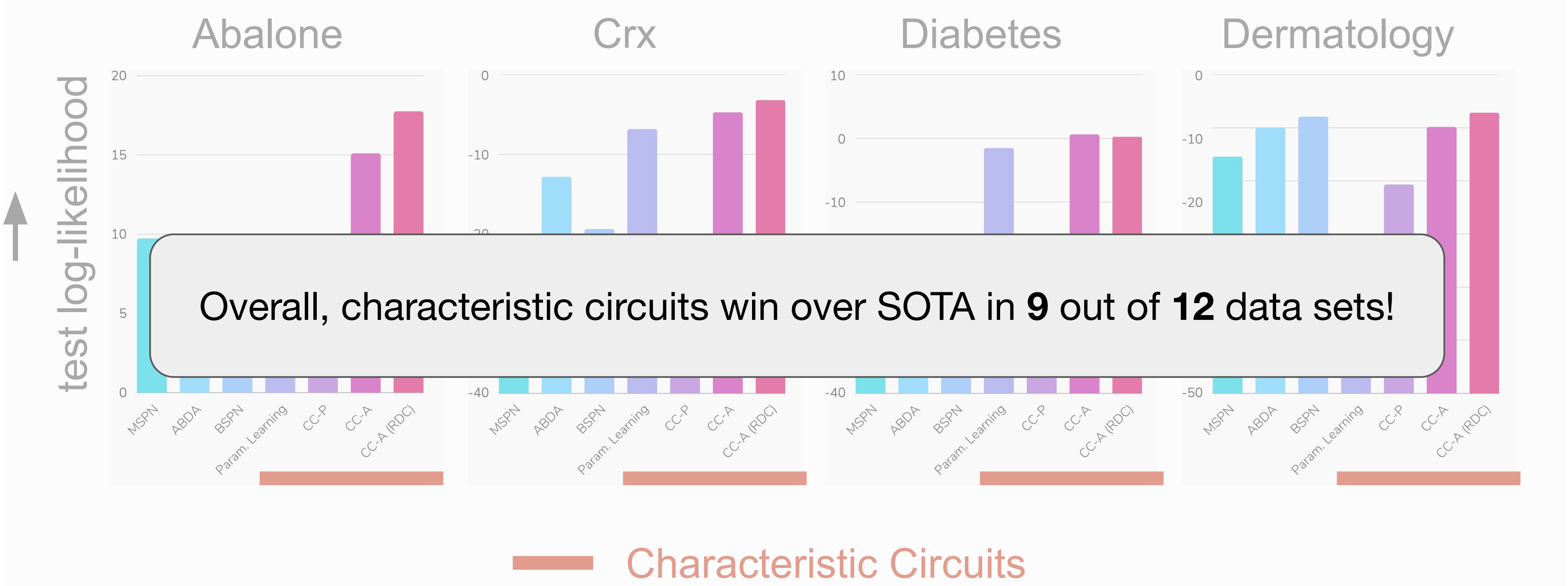
Characteristic Circuits

Better Density Estimators on Heterogeneous Data



Characteristic Circuits

Better Density Estimators on Heterogeneous Data



Overall, characteristic circuits win over SOTA in **9** out of **12** data sets!

TECHNISCHE UNIVERSITÄT DARMSTADT

Aalto University School of Science

Take-Away

Characteristic Circuits:

- a *unifying view* of hybrid models in the *spectral domain*
- *tractable* densities, marginals, and moments
- *parameter* and *structure* learning

Future directions:

- other queries from CCs, e.g. sampling
- advanced structure for parameter learning
- Characteristic Flows
- …

Thank you!

Questions?

Paper

Code

NEURAL INFORMATION PROCESSING SYSTEMS

Should we use it?

- Characteristic Circuits are better density estimators on heterogeneous data.
  Characteristic Circuits provide higher LL on 9 out of 12 of the data sets.

Data Set	Parameter Learning	MSPN	ABDA	BSPN	Structure Learning		
					CC-P	CC-A	CC-A^{RDC}
Abalone	3.06	9.73	2.22	3.92	4.27	15.10	**17.75**
Adult	-14.47	-44.07	-5.91	-4.62	-31.37	-7.76	**-1.43**
Australian	-5.59	-36.14	-16.44	-21.51	-30.29	-3.26	**-2.94**
Autism	-27.80	-39.20	-27.93	**-0.47**	-34.71	-17.52	-15.5
Breast	-20.39	-28.01	-25.48	-25.02	-54.75	-13.41	**-12.36**
Chess	-13.33	-13.01	-12.30	**-11.54**	-13.04	-13.04	-12.40
Crx	-6.82	-36.26	-12.82	-19.38	-32.63	-4.72	**-3.19**
Dermatology	-45.54	-27.71	-24.98	-23.95	-30.34	-24.92	**-23.58**
Diabetes	-1.49	-31.22	-17.48	-21.21	-23.01	**0.63**	0.27
German	-19.54	-26.05	-25.83	-26.76	-27.29	-15.24	**-15.02**
Student	-33.13	-30.18	-28.73	-29.51	-31.59	-27.92	**-26.99**
Wine	0.32	-0.13	-10.12	-8.62	-6.92	13.34	**13.36**
# best	0	0	0	2	0	1	9

[1] Alejandro Molina et al., "*Mixed sum-product networks: A deep architecture for hybrid domains*", AAAI, 2018.
[2] Antonio Vergari et al., "*Automatic bayesian density analysis*", AAAI, 2019.
[3] Martin Trapp et al., "*Bayesian learning of sum-product networks*", NeurIPS, 2019.

TECHNISCHE UNIVERSITÄT DARMSTADT

A! Aalto University School of Science

NEURAL INFORMATION PROCESSING SYSTEMS

Should we use it?

- Characteristic Circuits are better density estimators on heterogeneous data.
  Characteristic Circuits provide higher LL on 9 out of 12 of the data sets.

Data Set	Parameter Learning	MSPN	ABDA	BSPN	Structure Learning		
					CC-P	CC-A	CC-A^{RDC}
Abalone	3.06	9.73	2.22	3.92	4.27	15.10	**17.75**
Adult	-14.47	-44.07	-5.91	-4.62	-31.37	-7.76	**-1.43**
Australian	-5.59	-36.14	-16.44	-21.51	-30.29	-3.26	**-2.94**
Autism	-27.80	-39.20	-27.93	**-0.47**	-34.71	-17.52	-15.5
Breast	-20.39	-28.01	-25.48	-25.02	-54.75	-13.41	**-12.36**
Chess	-13.33	-13.01	-12.30	**-11.54**	-13.04	-13.04	-12.40
Crx	-6.82	-36.26	-12.82	-19.38	-32.63	-4.72	**-3.19**
Dermatology	-45.54	-27.71	-24.98	-23.95	-30.34	-24.92	**-23.58**
Diabetes	-1.49	-31.22	-17.48	-21.21	-23.01	**0.63**	0.27
German	-19.54	-26.05	-25.83	-26.76	-27.29	-15.24	**-15.02**
Student	-33.13	-30.18	-28.73	-29.51	-31.59	-27.92	**-26.99**
Wine	0.32	-0.13	-10.12	-8.62	-6.92	13.34	**13.36**
# best	0	0	0	2	0	1	9

[1] Alejandro Molina et al., "*Mixed sum-product networks: A deep architecture for hybrid domains*", AAAI, 2018.
[2] Antonio Vergari et al., "*Automatic bayesian density analysis*", AAAI, 2019.
[3] Martin Trapp et al., "*Bayesian learning of sum-product networks*", NeurIPS, 2019.

TECHNISCHE UNIVERSITÄT DARMSTADT

Aalto University School of Science

NEURAL INFORMATION PROCESSING SYSTEMS

Should we use it?

- Characteristic Circuits are better density estimators on heterogeneous data.
  Characteristic Circuits provide higher LL on 9 out of 12 of the data sets.

Data Set	Parameter Learning	MSPN	ABDA	BSPN	Structure Learning		
					CC-P	CC-A	CC-A^{RDC}
Abalone	3.06	9.73	2.22	3.92	4.27	15.10	**17.75**
Adult	-14.47	-44.07	-5.91	-4.62	-31.37	-7.76	**-1.43**
Australian	-5.59	-36.14	-16.44	-21.51	-30.29	-3.26	**-2.94**
Autism	-27.80	-39.20	-27.93	**-0.47**	-34.71	-17.52	-15.5
Breast	-20.39	-28.01	-25.48	-25.02	-54.75	-13.41	**-12.36**
Chess	-13.33	-13.01	-12.30	**-11.54**	-13.04	-13.04	-12.40
Crx	-6.82	-36.26	-12.82	-19.38	-32.63	-4.72	**-3.19**
Dermatology	-45.54	-27.71	-24.98	-23.95	-30.34	-24.92	**-23.58**
Diabetes	-1.49	-31.22	-17.48	-21.21	-23.01	**0.63**	0.27
German	-19.54	-26.05	-25.83	-26.76	-27.29	-15.24	**-15.02**
Student	-33.13	-30.18	-28.73	-29.51	-31.59	-27.92	**-26.99**
Wine	0.32	-0.13	-10.12	-8.62	-6.92	13.34	**13.36**
# best	0	0	0	2	0	1	9

[1] Alejandro Molina et al., “*Mixed sum-product networks: A deep architecture for hybrid domains*”, AAAI, 2018.
[2] Antonio Vergari et al., “*Automatic bayesian density analysis*”, AAAI, 2019.
[3] Martin Trapp et al., “*Bayesian learning of sum-product networks*”, NeurIPS, 2019.

TECHNISCHE UNIVERSITÄT DARMSTADT

Aalto University School of Science

Should we use it?

- Characteristic Circuits are better density estimators on heterogeneous data.
  Characteristic Circuits provide higher LL on 9 out of 12 of the data sets.

Data Set	Parameter Learning	MSPN	ABDA	BSPN	Structure Learning		
					CC-P	CC-A	CC-A^{RDC}
Abalone	3.06	9.73	2.22	3.92	4.27	15.10	**17.75**
Adult	-14.47	-44.07	-5.91	-4.62	-31.37	-7.76	**-1.43**
Australian	-5.59	-36.14	-16.44	-21.51	-30.29	-3.26	**-2.94**
Autism	-27.80	-39.20	-27.93	**-0.47**	-34.71	-17.52	-15.5
Breast	-20.39	-28.01	-25.48	-25.02	-54.75	-13.41	**-12.36**
Chess	-13.33	-13.01	-12.30	**-11.54**	-13.04	-13.04	-12.40
Crx	-6.82	-36.26	-12.82	-19.38	-32.63	-4.72	**-3.19**
Dermatology	-45.54	-27.71	-24.98	-23.95	-30.34	-24.92	**-23.58**
Diabetes	-1.49	-31.22	-17.48	-21.21	-23.01	**0.63**	0.27
German	-19.54	-26.05	-25.83	-26.76	-27.29	-15.24	**-15.02**
Student	-33.13	-30.18	-28.73	-29.51	-31.59	-27.92	**-26.99**
Wine	0.32	-0.13	-10.12	-8.62	-6.92	13.34	**13.36**
# best	0	0	0	2	0	1	9

[1] Alejandro Molina et al., "*Mixed sum-product networks: A deep architecture for hybrid domains*", AAAI, 2018.
[2] Antonio Vergari et al., "*Automatic bayesian density analysis*", AAAI, 2019.
[3] Martin Trapp et al., "*Bayesian learning of sum-product networks*", NeurIPS, 2019.